Select before you Collect

*Bart Jacobs**

Grootschalige gegevensopslag, zoals voorgesteld bij recente maatregelen ter bestrijding van terrorisme, introduceert nieuwe gevaren, zoals misbruik. Hierdoor, en door de manieren waarop burgers zich kunnen beschermen, is effectiviteit niet vanzelfsprekend.

In ons dagelijkse leven laten wij, bewust of onbewust, een grote hoeveelheid digitale sporen na, bijvoorbeeld via telefonie, e-mail, websurfen, betalen, reizen, digitale tv, enzovoorts. Deze gegevens worden kortere of langere tijd opgeslagen in digitale archieven. De technische regulering van toegang tot zulke, vaak gevoelige, gegevens is een onderdeel van het vakgebied computerbeveiliging. De juridische regulering vindt voornamelijk plaats via de Wet Bescherming Persoonsgegevens (WBP) en de Wet Computer Criminaliteit (WCC).

Het afgelopen jaar hebben de Europese ministers van Binnenlandse Zaken en van Justitie, verenigd in de zogeheten JBZ-raad,[1] plannen besproken om de aanbieders van communicatiediensten (telefoon en internet) te verplichten om de verkeersgegevens van alle 450 miljoen Europese burgers voor langere tijd op te slaan. De details van deze plannen veranderen nog voortdurend, maar de nadruk lijkt te liggen op de opslag van adresinformatie voor e-mail en telefoon (inclusief locatiegegevens voor GSM). Het opslaan van adressen van alle bezochte webpagina's behoort echter ook tot de opties. Tegelijkertijd hebben de Nederlandse JBZ-ministers in hun 'terreurbrief'[2] aan de Tweede Kamer het voornemen kenbaar gemaakt om de biometrische gegevens die in het nieuwe paspoort opgeslagen zullen worden ook in een centrale databank op te nemen, die toegankelijk zal zijn voor 'on line verificatie'.[3] De terreurbrief zegt: 'De ontwikkeling van deze informatie-infrastructuur draagt bij aan de intensivering van de samenwerking op Europees terrein en levert een bijdrage aan de effectiviteit van de uitvoering van de identificatieplicht.' Het gaat hierbij om een fundamentele beleidswijziging die ingaat tegen de eerdere interpretatie van minister van Boxtel, waarbij biometrische gegevens alleen gebruikt zouden worden om de eventuele *match* tussen persoon en paspoort vast te stellen, hetgeen geen centrale opslag vereist. Dit artikel zal nader ingaan op de vooral technische risico's die verbonden zijn aan grootschalige opslag van privacy-gevoelige gegevens.

Op de achtergrond speelt de discussie over privacy en veiligheid. Daarbij is de nuance in de politiek en media soms ver te zoeken

Op de achtergrond speelt de discussie over privacy en veiligheid. Daarbij is de nuance in de politiek en media soms ver te zoeken, getuige uitspraken als: 'privacy is de schuilplaats van het kwaad', 'wie niets te verbergen heeft, heeft ook niets te vrezen', 'we moeten privacy opofferen voor veiligheid', en soms zelfs: 'wil jij dan verantwoordelijk zijn voor de volgende aanslag?'. Die laatste uitspraak is natuurlijk uiterst effectief om iedere rationele discussie om zeep te helpen. Op deze plaats wordt ervan uitgegaan dat privacy intrinsieke waarde heeft, zonder dieper op deze materie zelf in te gaan.[4]

Zoals gezegd, het vakgebied computerbeveiliging richt zich op de regulering van toegang tot gevoelige digitale gegevens. Voorbeelden zijn militaire, medische of commerciële gegevens. Privacy is daarbij vooral een technische uitdaging waarbij het er om gaat enkel de volstrekt noodzakelijke gegevens te identificeren, versleuteld te transporteren en gecontroleerd te verwerken. Een goed voorbeeld is het eerdere plan van Roel Pieper voor re-

* Prof.dr. B. Jacobs is hoogleraar Beveiliging en Correctheid van computerprogrammatuur en verbonden aan het Institute for Computing and Information Sciences aan de Radboud Universiteit Nijmegen. Tevens is hij als hoogleraar verbonden aan de Faculteit Wiskunde en Computerwetenschap van de Technische Universiteit Eindhoven. Dit artikel is gebaseerd op een aantal voordrachten onder dezelfde titel, waarvan één op het Govcert Symposium van 8 en 9 september 2005. De bijbehorende Engelstalige presentatie is beschikbaar op *www.cs.ru.nl/B.Jacobs/PAPERS/govcert05.pdf*. De titel is bewust niet vertaald.

1 Raad Justitie en Binnenlandse Zaken.

2 *Kamerstukken II* 5327519/05/NCTb van 24 januari 2005.

3 Zie ook *Kamerstukken II*, BPR2005/54982, brief van minister Pechtold op 18 april 2005.

4 De auteur beschouwt zichzelf als een brave burger die zich netjes aan de wet houdt en keurig belasting betaalt. Desondanks heeft hij veel te verbergen.

keningrijden.[5] Daarbij zou iedere auto uitgerust moeten worden met GPS, voor plaatsbepaling met satellieten, en GSM, voor maandelijkse rapportage voor berekening van de heffing. Daarvoor is het echter niet noodzakelijk dat de betrokken autoriteiten te weten komen waar iedere auto precies geweest is. In de plannen zou Nederland verdeeld worden in wegen met verschillende kleuren, corresponderend met bepaalde tarieven. De maandelijkse rapportage hoeft dan enkel door te geven hoeveel kilometer op welke kleur gereden is. Dat is de kern van de zaak. Desgewenst kan een *override* knop aan het systeem toegevoegd worden waarbij opsporings- en inlichtingendiensten selectief toegang kunnen krijgen tot meer gedetailleerde locatiegegevens, op basis van autorisatie van hogere niveaus en met uitgebreide *logging* voor verantwoording achteraf.

De trend die zich af lijkt te tekenen is de volgende: traditioneel vindt eerst selectie plaats, bijvoorbeeld van een verdachte op basis van een redelijk vermoeden. Vervolgens kan pas overgegaan worden tot eventuele aantasting van de privacy van de verdachte, zoals door telefoontaps of huiszoekingen. Met de toegenomen informatietechnische mogelijkheden is een omdraaiing van selecteren en verzamelen binnen bereik gekomen: eerst van iedereen gegevens verzamelen en pas daarna een selectie maken van wie mogelijk gegevens gebruikt gaan worden voor nader (strafrechtelijk) onderzoek. Zo wordt in principe iedereen als potentiële verdachte gezien. Het komt mij voor dat hier een fundamenteel aspect van onze rechtsstaat in het gedrang komt en dat een geruisloze omdraaiing van selecteren en verzamelen niet gewenst is.

Vanzelfsprekend maakt 'eerst alles verzamelen' het opsporingswerk makkelijker, zeker in internationaal verband. Het maakt de opspoorders echter ook luier, omdat het niet langer nodig is in een vroeg stadium tot selectie over te gaan. Daarnaast biedt het mogelijkheden tot *datamining* en profilering, waarover later meer. Het is hierbij relevant op te merken dat de aanstaande Wet Computercriminaliteit II zogenaamde bevriezingsbevelen kent, waarbij communicatieaanbieders in een vroeg stadium opgedragen kan worden gegevens van geselecteerde partijen vast te houden. Ondanks dat deze bevriezingsbevelen nog helemaal niet ingevoerd en uitgeprobeerd zijn wordt al aan verdergaande maatregelen gewerkt. Relevante aspecten uit de discussie over grootschalige opslag van privacygevoelige gegevens zijn dus: de omdraaiing van selecteren en verzamelen, effectiviteit, proportionaliteit en risico's. De nadruk ligt in de rest van dit verhaal voornamelijk op het laatste punt.

1 Risico's gegevensopslag

Grote databanken van privacygevoelige gegevens kennen eigen risico's, zeker wanneer ze gebruikt worden voor profilering en daarop gebaseerde besluitvorming, zoals uitsluiting van toegang tot vliegtuigen (via no-fly lists), bus en trein (makkelijker via nieuwe OV-chipkaart), of ook verzekeringen of leningen. Naïef vertrouwen in de correctheid van gegevens in databanken is niet gerechtvaardigd. Bekend is het geval van de Amerikaanse senator Ted Kennedy die meerdere malen niet aan boord mocht voor een binnenlandse vlucht op basis van een persoonsverwisseling. Kennedy heeft verschillende keren contact op moeten nemen met minister Tom Ridge van *Homeland Security* om de fout hersteld te krijgen. Wat moeten u en ik in zo'n situatie? Onlangs werd bekend dat zelfs een 1-jarige baby op de no-fly list stond. Dichter bij huis zijn politie-invallen op het verkeerde adres niet onbekend.

Naïef vertrouwen in de correctheid van gegevens in databanken is niet gerechtvaardigd

Een ander risico is misbruik van gegevens. Dit risico is groter naarmate de toegang laagdrempeliger is. Bij de politie komt het opvragen van kentekeninformatie bij privé-aankoop van een tweedehands auto geregeld voor. Het rapport *Integriteit in het dagelijks politiewerk* zegt:[6] 'Min of meer algemeen geaccepteerd onder politiemensen lijkt het bevragen van politiële informatiesystemen voor privé-gebruik.' Onlangs is beweerd dat vanuit de Hofstadgroep via een apotheek gepoogd is adresgegevens te krijgen van politici. Grote databanken bij providers zullen een goudmijn vormen voor allerlei kwaadaardige activiteiten. Er mag dus verwacht worden dat de vele beheerders kwetsbaar zullen zijn voor benadering vanuit criminele milieus. Het gaat dus niet alleen om vrijwillig, maar ook om onvrijwillig misbruik door degenen die rechtmatig toegang hebben.

5 Zie het rapport *Mobimiles. Bewust op weg.* Een rapport van prof.ir. Roel Pieper in opdracht van minister Netelenbos van Verkeer en Waterstaat, Bloemendaal: 10 april 2001.

6 J. Naeyé, L. Huberts, C. van Zweden, V. Busato en B. Berger, *Integriteit in het dagelijks politiewerk. Meningen en ervaringen van politiemensen,* zie speciaal sectie 2.5. Zeist: Kerckebosch 2004.

Ingrijpend is het wanneer iemand informatie uit zo'n databank, bijvoorbeeld over een creditcard of toegangspas, misbruikt om zich als een ander voor te doen. Dergelijke digitale identiteitsroof wordt gezien als een van de snelst groeiende vormen van misdaad. Een slachtoffer van identiteitsroof kan bijvoorbeeld onterecht hoge rekeningen krijgen, vals beschuldigd worden of uitgesloten worden van bepaalde rechten. Wanneer uw pincode bekend raakt kunt u de bank om een nieuwe vragen. Bij biometrische gegevens, zoals een vingerafdruk of een irisscan, is vervanging onmogelijk. Zulke gegevens dienen dus extra goed beschermd te worden.

Wanneer uw pincode bekend raakt kunt u de bank om een nieuwe vragen. Bij biometrische gegevens, zoals een vingerafdruk of een irisscan, is vervanging onmogelijk

De beveiliging van databanken tegen inbraak of hacken is niet eenvoudig en vraagt constante aandacht, zeker wanneer er sprake is van externe toegang, bijvoorbeeld via het internet. Onlangs is gebleken hoe slecht het gesteld is met de beveiliging van digitale medische gegevens.[7] Op dit moment geeft de overheid providers slechts een zeer geringe vergoeding voor de extra kosten voor de opgedragen opslag. Het is dus geen onredelijke voorspelling dat providers geen maximale aandacht zullen besteden aan de beveiliging van verkeersgegevens. Het kost ze immers alleen maar geld. Vroeg of laat zullen zulke databanken omvallen en komen de daarin opgeslagen privacy-gevoelige gegevens op straat te liggen. Daar zijn betrokkenen het over eens. Het wordt dan bijvoorbeeld mogelijk om van individuen (zoals politici) systematisch vast te stellen of ze wel eens in hoerenbuurten komen, op basis van de locatiegegevens van hun gsm.

2 Profilering

Op basis van grote databanken vindt in toenemende mate profilering plaats: het opstellen van een individueel profiel uitgaande van de beschikbare gegevens, leidend tot een virtuele identiteit. In de commerciële wereld is dit zeer gebruikelijk. In Europa leggen de relatief strenge privacywetten beperkingen op aan het gebruik van zulke profielen, maar in de Verenigde Staten zijn er onbegrensde mogelijkheden.[8] Oorspronkelijk was dit vooral terrein voor private ondernemingen, met ChoicePoint en LexisNexis (dochter van Reed-Elsevier) als grote spelers, maar sinds de aanslagen van 9/11 is er een intensieve samenwerking met de eigen overheid.

Profilering wordt in de commerciële sector graag gebruikt voor gerichte marketing. Dit scheelt kosten en kan potentiële klanten soms verleiden omdat ze zich persoonlijk aangesproken voelen. Marketeers zijn echter niet direct in u persoonlijk geïnteresseerd; *They want to learn about people* like *you.*[9] *Eventuele fouten zijn niet rampzalig. Dat is anders bij profielen in de beveiligingssector, zowel bij false positives* (terrorist mag wel aan boord) als bij *false negatives* (baby mag niet aan boord). Een serieus probleem is dat de beoordelingscriteria meestal verborgen zijn en misschien wel van bedenkelijke kwaliteit of aard. Kun je maar beter geen halal-maaltijd bestellen op een vlucht naar de Verenigde Staten? In dit soort situaties wil je geen *bad profile* krijgen, omdat het deuren voor je sluit. Dit kan leiden tot conformistische neigingen uit angst. Het lijkt mij dat deze ontwikkelingen richting vergaand gebruik van profilering niet zonder regulering en toezicht kunnen.

Een overheid die haar burgers wil kunnen identificeren dient er tegelijkertijd voor te zorgen dat de mogelijkheden voor identiteitsroof minimaal zijn

Verder is het de vraag of profilering wel zo effectief is, bijvoorbeeld bij terrorismebestrijding.[10] Stel we hebben een extreem goed profiel met een (onrealistische) foutmarge van 1%. Hoeveel terroristen zijn er? Dat weten we natuurlijk niet precies, maar laten het er eens 1 op de miljoen zijn. Dan moet je dus 10.000 mensen uit de rij halen om die ene (mede) te selecteren, waarbij je hoopt dat het je vervolgens lukt de gezochte terrorist met andere

7 Zie: K. Spaink, *Medische Geheimen,* Amsterdam: Nijgh & Van Ditmar 2005.
8 Zie bijv. Robert O'Harrow, *No Place to Hide,* Detroit: Free Press 2005.
9 L. Lessig, *Code and Other Laws of Cyberspace,* New York: Basic books 1999.
10 Zie ook: B. Schneier, *Beyond Fear. Thinking Sensibly About Security in an Uncertain World,* New York: Springer-Verlag 2003.

middelen te identificeren in de overgebleven groep van 10.000. Dit is niet werkbaar. In lijn met de *select before you collect* gedachte steunen inlichtingendiensten eerst vooral op traditionele *eyes and ears (humint)* voor de selectie van bronnen van mogelijk gevaar en gebruiken ze vervolgens hun uitgebreide bevoegdheden voor data-verzameling (*sigint)* voor verdergaande focus op deze bronnen.

De risico's waar ik me concreet zorgen over maak zijn: (1) persoonlijke gegevens die op straat komen te liggen door slechte beveiliging; (2) op een of andere onduidelijke manier een *bad profile* krijgen; (3) identiteitsroof. Bijzonder problematisch is de mogelijke combinatie van deze drie punten. Door slordige omgang met mijn gegevens kan iemand anders zich als mij voordoen, mij veel schade berokkenen en er ook nog voor zorgen dat ik een *bad profile* krijg waardoor er niets meer voor mij werkt. Het lijkt mij voor de hand liggend dat een samenleving de risico's op dergelijke scenario's dient te minimaliseren. Bescherming van privacy is daarmee meer dan een persoonlijke aangelegenheid. Een overheid die haar burgers wil kunnen identificeren dient er tegelijkertijd voor te zorgen dat de mogelijkheden voor identiteitsroof minimaal zijn. Grootschalige gegevensverzameling staat daarmee op zeer gespannen voet.

Individueel kunnen burgers zich in zekere mate beschermen tegenover weetgierige overheden en bedrijven

3 Bescherming

Bescherming dient allereerst te komen via adequate wetgeving en regulering, vanuit het besef van risico's van gegevensopslag. Artikel 10 van de Grondwet is daar al op gericht. Uitwerking en aanpassing aan nieuwe technieken en methoden (zoals grootschalige profilering) is een continue opgave. In de VS moeten in verschillende staten bij compromittering van gegevens uit databanken alle betrokkenen sinds kort ingelicht worden.

Individueel kunnen burgers zich in zekere mate beschermen tegenover weetgierige overheden en bedrijven. Het basisprincipe is natuurlijk: nooit meer informatie weggeven dan strikt noodzakelijk is. Bewust ruis creëren kan ook, bijvoorbeeld door het onderling uitwisselen van spaarkaarten of het opgeven van onjuiste gegevens (zover mogelijk en wettig). Daarnaast zijn er technische beschermingsmogelijkheden, zoals het gebruik van versleutelde e-mail (via het zogenaamde PGP-systeem) of anoniem websurfen (via het zogenaamde Tor-systeem). De daarvoor benodigde software is gratis beschikbaar en installatie en gebruik zijn niet onoverkomelijk moeilijk. Verder kan communicatie ook via bijvoorbeeld *multiplayer games* verlopen. De aanwezigheid van dergelijke middelen relativeert de effectiviteit van geplande wetgeving voor opslag van verkeersgegevens. Alleen de allerdomste criminelen laten zich er mee in de kaart kijken.

Naarmate burgers dwarser worden heeft data surveillance minder zin

Een interessant alternatief is *privacy flooding*. Dit is een vlucht vooruit, waarbij de gevoelige gegevens bewust geopenbaard worden, maar op zodanige wijze dat ze hun zin verliezen. Te denken valt aan een scenario waarbij veel Nederlanders hun vingerafdruk op het web zetten. Dit maakt biometrische identificatie problematisch: wanneer mijn vingerafdruk dan op een plaats delict aangetroffen wordt kan ik altijd beweren dat iemand anders die van het web gehaald heeft en in siliconen nagemaakt. Terzijde: bij DNA is dit moeilijker, maar uiteindelijk niet uitgesloten. Vergelijkbaar zou ik een computerprogramma kunnen schrijven en verspreiden dat willekeurig webpagina's opvraagt (maar niet in mijn browser laat zien). De pagina's waar ik werkelijk in geïnteresseerd ben zijn dan moeilijk te identificeren in de overvloed die bij mijn provider voorbij komt. Deze observaties verschaffen een nieuw argument voor terughoudendheid: naarmate burgers dwarser worden heeft *data surveillance* minder zin.[11]

Door het voortschrijden van technische mogelijkheden en door het veranderen van maatschappelijke omstandigheden blijft het noodzakelijk de regels voor opslag en verwerking van gegevens voortdurend bij te stellen. Dit is duidelijk een gebied waar niet alles wat technisch mogelijk is ook gedaan moet worden. Terroristen dagen ons uit om juist die waarden te ondermijnen die we claimen te verdedigen. Laten we ze hun zin niet geven!

11 Een meer agressieve tactiek is om de leiders en initatiefnemers van *data surveillance* persoonlijk in hun privacy aan te tasten. Dit overkwam John Poindexter als leider van het inmiddels beëindigde *Total Information Awareness (TIA)* programma van het Amerikaanse ministerie van Defensie. Zie: P.R. Keefe, *Chatter. Dispatches from the secret world of global eavesdropping*, New York: Random House 2005.